

Molecular Speciation Effect on Docking and Drug Design. A Computational Study for Mangiferin, a Carbohydrate-Polyphenol Bioconjugate as a Test Case

Berenice Gómez-Zaleta,^{1,2*} Claudia Haydée González-De la Rosa,¹ Gerardo Pérez-Hernández,^{1,2} Hiram I. Beltrán,¹ Felipe Aparicio,^{1,2} Alberto Rojas-Hernández,² and Arturo Rojo-Domínguez^{1,2*}

¹Departamento de Ciencias Naturales. Universidad Autónoma Metropolitana. Unidad Cuajimalpa. Pedro Antonio de los Santos 84. Col. San Miguel Chapultepec. 11850 México, D.F. E-mail: ard@xanum.uam.mx.

²Departamento de Química. Universidad Autónoma Metropolitana. Unidad Iztapalapa. San Rafael Atlixco 186. Col. Vicentina, 09340 México, D.F.

Recibido el 3 de octubre del 2007; aceptado el 21 de febrero del 2008

Abstract. A study to evaluate the effect of molecular speciation considering methodologies to assign partial charges and conformational search processes for a docking test was made with mangiferin (MGF). This compound was selected as a model to explore speciation effects on drug design due to the speciation studies previously performed, and because it is a bioconjugate containing carbohydrate and polyphenolic xanthonoid groups, both moieties important as potential-drug candidates. PEOE (Partial Equalization of Orbital Electronegativity) resulted the best method to assign partial charges, with a good compromise between precision and computational cost, among different Classical Molecular Force Fields and Quantum Mechanics methods that were compared with Density Functional Theory calculations as the reference methodology. The number of conformations in energy minima showed to be extremely dependent upon partial charge assignment, as well as their geometry. In docking simulations of MGF on albumin drug-site 1, it was showed the relevance of choosing the properly expected chemical species for the pH value of interest since neutral MGF or deprotonated at the hydroxyl group on position 1 results in orientations significantly different from those predicted for the species deprotonated at the hydroxyl group on position 6, which is the predominant deprotonation site in accordance with the speciation study. The first two species present a tendency to expose the carbohydrate region to solvent occupying the same region in the binding site, while the molecule deprotonated in position 6 exhibits a preference for a different region of the site with its xanthonoid moiety exposed. Carbohydrate-polyphenol bioconjugates, such as MGF combine two types of bioactive molecules being both important as leaders for drug design.

Keywords: Xanthonoids, conformational search, molecular recognition, drug design, speciation.

Resumen. Se realizó un estudio para evaluar la importancia de la especiación química para asignar cargas parciales a moléculas mediante diversas metodologías y su aplicación en procesos de búsqueda conformacional para pruebas de acoplamiento molecular proteína-ligando. Estos efectos impactan en el diseño de fármacos y como caso particular se ha tomado a la mangiferina (MGF) como molécula modelo del tipo bioconjugado, tanto de glicósidos como de xantonoides, ya que ambos tipos de moléculas tienen potencial farmacológico. Las cargas parciales obtenidas por la teoría de funcionales de la densidad se utilizaron como referencia para comparar con otros métodos de asignación basados en campos de fuerza moleculares o en mecánica cuántica. El método PEOE (siglas en inglés de Equilibración Parcial de Electronegatividades Orbitales) fue el que tuvo mejores resultados, con un buen compromiso entre bajo costo computacional y precisión. Por su parte, tanto el número de conformaciones de baja energía como la geometría de cada una de ellas demostraron ser extremadamente dependientes del esquema de asignación de carga parcial. En las simulaciones de acoplamiento molecular para la MGF en el sitio 1 de albúmina se demostró la relevancia de elegir adecuadamente la especie química predominante en el valor de pH de interés dado que la molécula neutra de MGF, así como la molécula con el grupo hidroxilo desprotonado en posición 1 dan como resultado orientaciones significativamente diferentes de aquellas cuando la molécula se encuentra desprotonada en el grupo hidroxilo de la posición 6, la cual sería la especie predominante de acuerdo al estudio previo de especiación. Las primeras dos moléculas presentan una tendencia a ocupar el mismo lugar en el sitio de unión y exponer al solvente la región del carbohidrato mientras que la molécula desprotonada en la posición 6 muestra una tendencia a ocupar un lugar diferente en el sitio de unión de la albúmina y a exponer al solvente el grupo xantonoide. La MGF es un bioconjugado formado por un xantonoide y un carbohidrato, y siendo ambas partes constituyentes de moléculas bioactivas hace que nuestra molécula en estudio sea un buen caso modelo para la búsqueda de compuestos líder en el diseño de fármacos.

Palabras clave: Xantonoides, búsqueda conformacional, reconocimiento molecular, diseño de fármacos, especiación química.

Introduction

Computer assisted drug design has proved its utility during the last decades, for some reviews see references [1-5]. One of its approaches consists in modeling the structure of an optimal receptor-ligand complex by testing different orientations for a set of small organic molecules, with pharmacological potential, into a presumed binding site in the receptor, generally a protein. This method known as docking, and has been subject

of recent publications such as [6-10]. To decide for the best pose of a particular ligand, a numerical score is computed; the best one is selected, and compared with those from the other ligands. This score frequently rely on inter- and intra-molecular atomic interactions being electrostatics one of the main contributors to them. Consequently, it is of great importance to assign partial-charges accurately on the atoms of the ligand. In this work we explore a test case to evaluate the effect of different atomic-charge assignments and distinct protonation states of mangiferin (Scheme 1) as ligand on docking studies. MGF

has been taken as a bioconjugate representative member, with carbohydrate and xanthonoid moieties that made it a potential lead for drug design. Pursuing for an optional methodology with less computational cost but good enough accuracy, Density Functional Theory (DFT) was chosen as the most accurate method to be compared with, in order to evaluate the different approaches for atomic-charge assignments. During molecular simulations, and particularly in docking studies, the labile protons of ligands must be removed in order to give the molecule its correct protonation state. But this state is highly dependent on the experimental conditions, and we have selected MGF as a test case not only for being representative of drug-like compounds but also because our research group have made a previous speciation study on this molecule [11], thus allowing the analysis of the predictive results from this work.

Modeling complexes formed by a macromolecule (generally a protein) and a small organic molecule (a ligand) represent a challenge in terms of force field selection and charge assignment, since most of force fields specifically parameterized for macromolecules contain their own set of partial charges for atoms from proteins or nucleic acid, but not for the diversity of heteroatoms and their arrangements frequently found in xenobiotic small molecules. General purpose force fields, on the contrary, are designed to deal with a broader set of functional groups, but are not accurate enough for proteins. In this work we found that protein charges can be assigned by a specific force field but it is better to use PEOE for organic molecules. For the docking studies, as important as for the assignment of partial charges it is to know the distribution of chemical species in the system, which constitutes a speciation subject concern [12]. Chemical species are considered as "...the specific forms of an element defined as to isotopic composition, electronic or oxidation state, and/or complex or molecular structure..." accordingly to the IUPAC. In this sense, the main point is to identify the chemical species that will be present at the pH value of interest, 7.4 in this case, if they are electrically neutral or charged and where this charge would be located within the molecular framework. Mangiferin, as many polyphenolic structures, presents an interesting challenge since the determination of the equilibrium constants, and

more over their assignation in the structure, can not always be a simple task. In this work we first study different methods for charge assignation on MGF; then all docking studies of this molecule on a potential proteic receptor were made using the partial charge calculations from DFT results.

Polyphenols and among those xanthonoids, have been a biomolecular group of raising interest during the last three decades due to their documented properties as antimicrobial, antimalarian, and pharmacological agents [13]. Among polyphenols, mangiferin (D-glucopyranosyl-1,3,6,7 tetrahydroxyxanthen-9-one, Scheme 1) is naturally occurring in many higher plants from families as *Anarcardiaceae*, *Gentianaceae* and *Guttiferae* [14-16]; it has been widely studied as an antioxidant, antidiabetic as well as antiviral [17-24] compound, and it is also known its antituberculosis activity since 1975 [25].

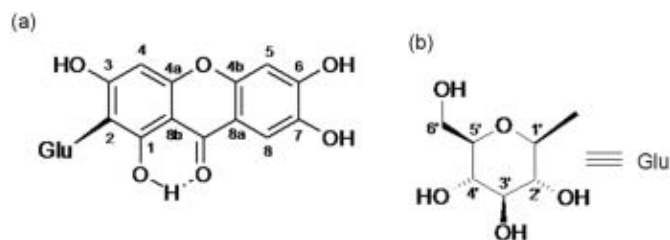
Mangiferin bioactivities have been related with radical scavenging [26] as well as inhibition of oxidative stress [27] and complex formation with Fe(III) [28] where knowledge regarding its fundamental chemical behavior could be helpful for a better understanding of its biological pathways. However, there is little information related to quantum mechanics for this molecule [29] and almost none about docking studies, essential to understand its mechanisms of biological action.

In this sense, speciation could have very important implications along its application in drug design. It is our purpose to show the significance of the speciation through possible effects on charge assignation and computer modeling of ligand-receptor complexes.

Results and Discussion

1. Partial Charge Assignments

During modeling of protein-ligand complexes, two interactions dominate, i) van der Waals, and ii) electrostatic. The first is modeled frequently by a classical 12-6 Lennard-Jones potential and the second by a coulombic term which is usually, by far, the most important. Both interactions are calculated as the sum of all atom-atom contacts formed by one atom of the ligand and one of the receptor. Docking is a procedure where different ligand conformations are tested, on a large set of orientations, in the receptor binding site. In order to choose the most suitable ligand conformation in the best receptor orientation, it is required a numeric score. This score is often highly influenced by electrostatic interactions and in consequence partial electric charges must be pre-assigned in both atoms from the counterparts, the ligand and the receptor molecules. It should be kept in mind that this is a rather oversimplified description of the docking process, where molecular flexibility is included only in the ligand by a set of low energy conformers, and solvation effects are indirectly incorporated in the hydrophobic contacts. Nevertheless, successes obtained in docking simulations indicate that a significant part of the effects are actually represented in the model. In this work the MGF molecule was



Scheme 1. Mangiferin structure in accordance with the reference [11]. (a) xanthonoid moiety, (b) glycoside or carbohydrate moiety. The oxygen and hydrogen atoms are numbered after the atoms they are attached to. Carbonyl and ether atoms in the xanthonoid moiety were respectively identified by C, O and OA.

selected as a model to explore two different problems in the partial charge assignment: i) the methodology used for it, and ii) the ionization state of the molecule, which is related to the chemical speciation study.

Three different schemes were used to assign the atomic partial-charge for the 48 atoms of neutral mangiferin. The first one is based on direct assignment of classical molecular force fields; the second one, is the widely used Partial Equalization of Orbital Electronegativities (PEOE) [30], and finally quantum mechanics methods were applied. It should be stated that a common practice is to employ a fast method in the charge assignment process instead of the computationally expensive quantum mechanics calculations, due to the vast number of molecules in chemolibraries of drug-like compounds (with hundreds of thousands of molecules) and in even larger libraries containing millions of compounds (e.g. Zinc Data base, [31]). Nevertheless, it was part of our goal to explore and evaluate different methodologies looking for the best trade-off between accuracy and computational cost, with the intent of selecting a cheaper alternative route, for being applied in the future to large sets of polyphenolic structures (such as flavonoid and xanthonoid derivatives) in an automatic computational procedure. Since the Density Functional Theory (DFT) calculations [32] are considered more accurate (but with greater computational cost) over classical molecular force fields as well as the PEOE, it was chosen here the DFT partial-charge assignment calculations as the most precise.

The results of charge assignment on neutral MGF are shown in Fig. 1. A general observation is that all methods employed tend to overestimate the magnitude of partial charge, but most of the times with the correct sign (exceptions include C2, C8a and C8b in xanthonoid, and none in carbohydrate). The highest individual differences are on the carbon atom of carbonyl group and on atoms of hydroxyl groups on both the xanthonoid and carbohydrate moieties, where the absolute value of the partial charge is overestimated by more than 0.4 atomic units. It can also be seen that, as expected, the most divergent values came from classical force fields. Similar results can be seen from the correspondent graphics for both deprotonated species (data not shown). For these later species, the mandatory value of -1.0 for the net sum of atomic charges is not achieved for methods based on parameterization of force fields (Table 1) except for MM94FF, so they were discarded hereafter for further studies. Moreover, the correlation coefficients obtained between charges from each method and those from DFT are shown in Table 1. It can be seen that the best global correlation is found with CHARMM and OPLS in classical force fields, a good performance is shown with PEOE and the best correlation of QM methods is found with AM1-BCC. It should be remarked that PEOE performed very well, resulting both, the best globally correlated with DFT data, and the method with the smallest individual overall differences.

Without further information, on a typical docking study, MGF would be considered as a neutral molecule, since its hydroxyl groups from the glycoside moiety or the phenol groups attached to the xanthone are expected to have a pKa

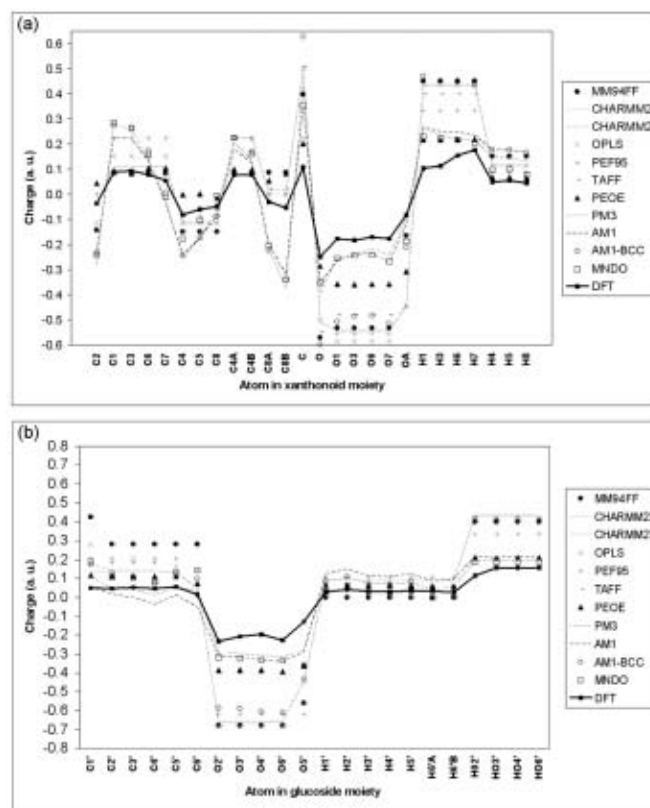


Fig. 1. Mangiferin atomic charge distribution calculated with different methods. (a) xanthonoid moiety; (b) glucoside moiety. Labels on horizontal axis identify atoms of MGF in accordance with Scheme 1.

value higher than the physiological pH when considered in isolation from the rest of the molecule. However, a more detailed approach to MGF (not feasible in a high throughput virtual screening study but viable for computational assays with a limited number of potential ligands) would estimate pKa values for the different chemical groups of this compound. For example, a prediction tool such as Advanced Chemistry Development [33] estimated a pKa value of 6.04 ± 0.2 for the hydrogen of the hydroxyl group bonded to C1, being thus predicted as the most acidic group, and values above 7.4 for the rest of the protons.

Nevertheless a further detail level, many times needed for advanced prototypes of pharmaceuticals and their bioinformatics studies, appears with a complete experimental speciation study. In the case of MGF its speciation study was previously reported by this research group [11] and it should be emphasized that such studies were comprised both experimental and theoretical in nature and are not practical to be performed even for small-sized chemolibraries. In that study we demonstrated that the most acidic proton of MGF is the one in the hydroxyl group bonded to C6, which is ionized at neutral pH, with a pKa value of 6.52 ± 0.06 while OH group bonded to C1 has a 12.10 ± 0.02 pKa value since it is stabilized by a hydrogen bond formed with the neighboring carbonyl oxygen. Actually,

Table 1. Sum of the partial charges for mangiferin obtained with different methods, and their correlation respect to DFT results. *MGF* = neutral mangiferin; *DP1* = mangiferin¹⁻ with deprotonation in position 1; *DP6* = mangiferin¹⁻ with deprotonation in position 6.

	Classical force fields						Quantum mechanical methods					
	MM94FF	CHARMM22	CHARMM27	OPLS	PEF95	TAFF	PEOE	PM3	AM1	AM1 -BCC	MNDO	DFT
Sum of charges on												
MGF	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0001	0.0001	-0.0001	0.0004
DP1	-1.0000	0.0000	0.0000	0.0000	0.0003	0.0000	-1.0003	-1.0009	-1.0003	-0.9999	-1.0002	-0.9992
DP6	-1.0000	0.0000	0.0000	0.0000	0.0003	0.0000	-1.0000	-1.0003	-0.9999	-0.9999	-1.0003	-0.9996
Correlation of charges												
MGF	0.94485	0.97115	0.97187	0.97078	0.94664	0.94309	0.94559	0.90461	0.92094	0.97208	0.92079	
DP1	0.93409	0.90995	0.91143	0.90899	0.88092	0.86348	0.93524	0.85702	0.87983	0.93634	0.87620	
DP6	0.92657	0.88872	0.88972	0.88806	0.86714	0.83961	0.94153	0.85322	0.87422	0.81686	0.86584	
Average of three correlations	0.93517	0.92327	0.92434	0.92261	0.89823	0.88206	0.94079	0.87162	0.89166	0.90843	0.88761	

from the four pKa values determined for the MGF molecule, it was the proton in position 1 the less acidic one. Indeed, it is interesting to explore the effect of the most acidic proton-misassignment on the conformational search and docking approaches since this type of studies are not usually performed during high throughput virtual screening of potential ligands. To analyze the speciation detail effect we carried out the same partial charge assignment as discussed for neutral MGF to the species deprotonated at position 1 or 6 (denoted as DP1 and DP6).

Global correlation between partial charges obtained by different methods and those resulting from DFT are shown in the lower rows of Table 1, along with the average of correlations for the three species of mangiferin: MGF, DP1 and DP6. The main divergences in the deprotonated species were again with oxygen atoms, hydrogen atoms attached to them and the carbon at carbonyl groups. Results on charge assignments for MGF have shown the failure to deal with ionized hydroxyl groups in all the types of the classic force fields employed, with the exception of MM94FF where the results show that it was the only force field analyzed which showed to be able to assign a unitary negative charge to deprotonated MGF. PEOE method again performed well due to the model construction that considers local polarization from the ionization of hydroxyl groups, through redistribution of fractional charges to first and second neighbors. This method is based on an iterative calculation that starts with all atoms with zero charge, and then a fraction of electron charge is transferred to all covalently bonded atoms, from the most electropositive to the most electronegative; the electronegativity of each atom is re-estimated to deal with their new partial charges and the pro-

cess continues until equilibrium is reached as all atoms have the same electronegativity. To obtain convergence, an exponentially decreasing function damps charge transference. Re-estimation of electronegativity is based on smooth functions fitted to previously reported calculations by quantum-mechanics on the electronegativity of each element carrying different amounts of partial charge. PEOE is a low cost calculation and seems to be the best method for charge assignment on both polyphenolic and glycolic moieties. For the following studies, the neutral and the two deprotonated forms of MGF were compared against DFT results using just the latter justified methodology along with AM1-BCC, since it had the best average correlation in semiempirical calculations. Due to parametrization of the MNDO, AM1 and PM3 methods, these semiempirical approximations tend to yield poor results for molecules where large charges are localized on atoms. The occupied orbitals of negatively charged species tend to expand in space due to electron repulsion. An analogous but opposite effect is noted for positively charged systems. These phenomena are accounted for in AM1-BCC and *ab initio* methods. Unlike the rest of the semiempirical methods used in this work, the AM1-BCC method was designed in order to produce atomic charges that emulate the HF/6-31G* electrostatic potential of a molecule. For this reason, we have a good correlation with the DFT results.

2. Conformer Search

In order to simulate ligand flexibility during docking studies, sets of conformers for each ligand to be tested are constructed; it is based on the observed fact that bound ligands adopt one of

those conformations when bounding their receptor. For testing the charge assignment effect on the conformational search of MGF, its different species described above were used: MGF, DP1 and DP6, as if they were three different molecules, each one with their own set of minimum energy conformations. The conformational search was not an easy task regardless the rigidity of the xanthonoid group and the known anchoring of the glucopyranosyl group. It is due to several degrees of conformational freedom on C2-C1' and C5'-C6' bonds and in all CO bonds close to hydroxyl groups (four in the xanthonoid and four in the glucopyranosyl moiety). A conservative estimation for possible conformers would search every 60° in each rotational bond from the ten independent ones, yielding $6^{10} > 60$ million structures. But each rotational bond is not independent and can bring atoms together producing a huge number of new conformations in energy minima.

A high throughput virtual screening can not spend a long time on each molecule looking for all its conformations situated in energy minima, and usually during the screening only random searches are performed detecting just a fraction of those geometries. Although there exists a variety of methodologies for conformational searches (for example, molecular dynamics, Monte Carlo algorithms and simulated annealing), in this work, considering that a full systematic search is beyond our scope, two intermediate procedures were combined to search for the highest possible number of conformers. Procedure A: xanthonoid and glucopyranosyl rings were considered rigid, then all conformers were searched systematically by setting each of the ten bonds with rotational freedom at angle values corresponding to pre-calculated energy minima. Procedure B: the search was performed emphasizing the rotation of the most important bond for xanthonoid-glucopyranosyl interactions, the C2-C1' bond, turning it every 30°, and the hydroxyl groups on C1 and C3 every 60°.

In both procedures all resultant conformations were fully energy minimized with MM94FF force field but maintaining the corresponding partial-charge assignments. Through the conformer minimization assays, two convergence limits had been used, 0.01 and 0.001 kcal·(mol Å)⁻¹, but it was chosen the first one due to the 30% decrement in computer time compared to the second with no significant difference in results.

To avoid redundancy, in each of the 30 searches performed, all geometries obtained were compared one to one superposing them to minimize the average distances between their equivalent atoms (calculated as root-mean square or RMS), and considering identical those with RMS lower than 0.1 Å. Finally, resultant conformers were ranked according to their energy. It should be mentioned that structures with energies higher than 4 kcal·mol⁻¹, respect to the lowest energy conformer, were generally discarded since the tensional energy needed to adopt such geometry is in detriment of the ligand-receptor affinity, where intermolecular contacts should overcompensate conformational tension.

For this analysis those minimized conformations with energies above 20 kcal mol⁻¹ were discarded, and for subsequent docking studies, those above 4 kcal mol⁻¹ were also

eliminated. The resulting number of conformers for the three studied species, five different charge assignments and both procedures were gathered and summarized in Table 2. Each column shows the number of conformers found. For example, it can be seen in PEOE column for MFG that procedure A generated 6625 initial conformers but after minimization only 60 different ones can be detected, 2 with energies above 20 kcal mol⁻¹ respect to the lowest energy geometry, and 58 below this threshold. From those 58 conformations, 41 have energies below 10 kcal mol⁻¹ and only 15 below 4 kcal mol⁻¹. A striking feature is that there is a huge difference in initial conformers in the two procedures but in most cases the number of final minimized conformers with low energy is rather similar with both procedures. A very interesting result is the number of initial conformers strongly depends on the charge assignment scheme, and more important, the number of final low energy conformers also depends on this. Further conformation analyses indicate that these do not coincide when using different charge assignment methods. Moreover, the lowest energy conformation found is not always the same neither for the different assignment methods nor for procedures A and B for the same assignment method. It is an indication of the potential energy surface complexity for this relatively simple molecule and the dependence of that surface on its variables such as partial charges. Also, it can be seen from figures in Table 2 that other generalizations are difficult to establish, again suggesting the system complexity. Differences in conformation and in the energy among different minima arise from rotation around the bond that binds the xanthonoid to the carbohydrate, which approaches or move away hydroxyls close to that bond in both moieties (OH groups supported on carbons 1, 3, 2' and 5', see Scheme 1) and also in the interactions between OH groups in the carbohydrate.

3. Docking Computational Experiments

As far as the authors know, no proteic receptor has been identified for MGF. But with the aim of understand the effect of charge assignment on conformational search and docking, we used serum albumin as a potential receptor. This protein is one of the most abundant carrier proteins in human plasma (≈600 μM) and plays an important role in transport and disposition of endogenous and exogenous ligands present in blood [34] including quercetin, a natural occurring flavonoid [35]. The structure of albumin bound to more than a dozen different drugs has being solved by crystallography [36] and different binding sites have been determined depending on the nature of the ligand counterparts as well as on the experimental conditions of the research. The most common of those sites is named drug-site 1. Although albumin might be considered a plasma carrier for MGF, there is no evidence of the participation of this protein in the therapeutic effects of MGF as a phytopharmaceutical [37], so in this work we considered the study of the MGF-albumin complex as a test case for exploring the chemical assignment effect of ionization state and partial charges on docking procedures, and we are not trying to

Table 2. Number of conformers generated for neutral and charged mangiferin species with different charge assignment methods and two searching procedures.

	MGF					DP1					DP6				
	MM94FF	PEOE	AM1-BCC	MNDO	DFT	MM94FF	PEOE	AM1-BCC	MNDO	DFT	MM94FF	PEOE	AM1-BCC	MNDO	DFT
NTCC_A	313	6,625	181	6,625	7,393	157	1,105	91	1,201	1,057	209	4,417	185	4,417	1,929
MC with*															
energy > 20	47	2	42	10	0	34	0	7	0	0	99	197	28	46	0
energy < 20	35	58	21	65	15	27	17	13	9	3	22	79	39	56	19
energy < 10	12	41	7	32	15	7	12	7	7	3	13	36	17	29	17
energy < 5	3	21	2	16	11	2	6	3	4	3	9	18	7	16	10
energy < 4	2	15	2	11	11	2	5	3	4	3	5	13	6	12	10
NTCC_B	7,957	35,191	11,026	8,209	32,473	2,263	15,595	1,306	16,951	31,549	12,793	86,581	11,317	88,237	81,295
MC with*															
energy > 20	2,280	16	4,489	12	0	327	6,914	163	68	0	7,670	4,320	3,858	664	0
energy < 20	83	58	34	63	39	103	19	42	59	30	57	129	65	76	47
energy < 10	21	41	12	31	33	46	19	15	19	19	18	52	30	31	30
energy < 5	6	21	5	8	13	20	6	3	5	8	6	22	16	8	13
energy < 4	5	15	4	5	11	19	5	2	4	8	5	16	11	5	12

NTCC = number of total conformations created by procedure A or B

MC = minimized non-redundant conformations

* energy figures in kcal/mol

assess the mechanism of the reported MGF medical activities. All conformers of MGF, DP1 and DP6 with charge assignments by DFT and with energy lower than 4 kcal mol⁻¹ were docked to albumin drug-site 1. The ten best scored poses for each of those molecules are shown on Figure 2. Although different orientations of the ligands can be found, two of them are the most observed; one with xanthonoid moiety (schematized as the arrow head) buried in the protein pocket, and the other with the carbohydrate occupying this bottom of the binding site. From these results, it was found that oxygen 1 (circled in Figure 2), either protonated or deprotonated in MGF and DP1 species, frequently interacts with the positive amino group on the side chain of LYS199, with the molecule on position showed by arrow A. Remarkably, DP6 species is oriented in a significantly different pose, being found by docking simulations preferably in the position represented by arrow B, with the deprotonated hydroxyl (squared 6) exposed to solvent.

The number of structures found on different orientations for the top 50 poses for MGF, DP1 and DP6 binding to albumin are shown in Table 3. This number of poses arises on one hand from having about 10 low energy conformations to deal with (see Table 2) combined with the idea of keeping the best five poses of each one. In the other hand, poses with scores out of the best 50 are quite low and with very similar values to allow real discrimination among them. Results are collected in groups of 10 residues. A' stands for an orientation similar to A but with the xanthonoid residue rotated 180° along the arrow axis, which means carbonyl and O1 pointing opposite to circled 1 in Figure 2. Carbohydrate moiety occupies the same region in A and A' orientations. The symbol -B corresponds to a B orientation but in opposite direction, which locates the carbohydrate close to the B arrow head and xanthonoid buried

in the lower part of Figure 2. Finally O stands for other orientations. It can be noticed that less than 5% of the results presented in Table 3 corresponds to this latter category, and that A' is also rarely found. For the neutral molecule a clear tendency is found: buried xanthonoid moiety is presented in the 80% of the analyzed cases (A + A' + B = 41 from 50 cases). A lower tendency to bury the xanthonoid group is found for DP1, with the lowest scores of the three species. In addition, DP1 presents the highest -B orientation incidence, due to interaction between its deprotonated oxygen and the basic side chains of Lys199 and His242 residues. However, the best scores were found for DP6, with a strong tendency to expose its deprotonated oxygen to solvent (orientation B) and to occupy the B and -B cleft. It is an interesting result since the DP6 species is proposed, in the speciation study [11], as the predominant one to be expected at pH = 7.4.

Methods

Molecular modeling, charge assignment, conformer searching, visualization and docking were performed with Molecular Operating Environment (MOE) package [38] version 2006.08, with default parameters unless otherwise stated. Energy minimizations were carried out with MM94FF force field until an RMS force lower than 0.01 or 0.001 kcal mol⁻¹ Å⁻¹ was obtained.

Charge assignments on electrically neutral MGF were calculated by six different classical force fields including MM94FF suited for small molecules [39-40], CHARMM 22 parameterized for proteins and nucleic acids [41], CHARMM27 parameterized for proteins and heme groups [42], OPLS

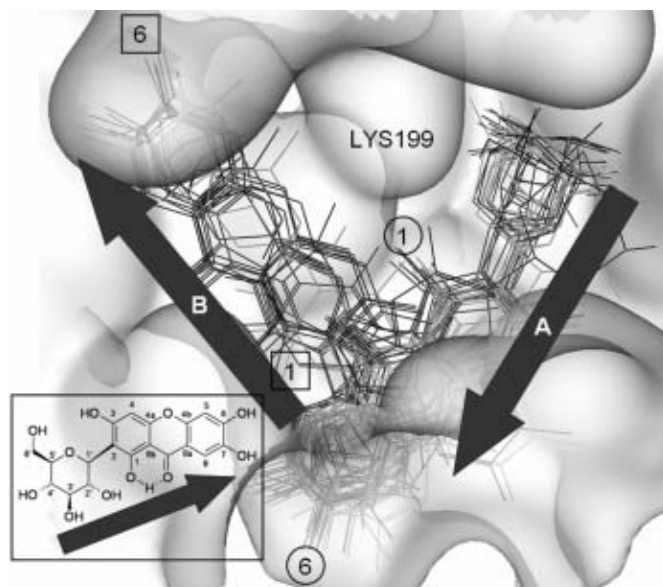


Fig. 2. Detail of the molecular surface of albumin around drug-site 1. Arrows A and B represent the main orientations of the ligand on optimal poses in the binding site, in the direction of carbohydrate to xanthonoid as depicted in the inset. Hydroxyl groups 1 and 6 of the ligand are identified by circled and squared numbers on orientations A and B, respectively. The top 10 scores are shown for each mangiferin species (docking scores obtained for the top 10 orientations range from -5.98 to -5.26 for MGF, -5.77 to -5.03 for DP1 and -6.35 to -5.32 for DP6).

designed for proteins and small organic molecules [43], PEF95 for carbohydrates [44] and TAFF parameterized for small molecules (adapted in MOE from Tripos all Atom Force Field [45] included in Sybyl program –www.tripos.com). Amber force field was not used since it uses PEOE assignment for organic molecules and gave no additional information. Engh-Huber force field [46] was discarded since it uses a united atom scheme (which implies non polar hydrogens fused to the carbon atoms that support them) opposite to the rest of the methods which used all-atom description for molecules. Quantum mechanics methods included partial charges calculated from the PM3, AM1 and MNDO semi-empirical Hamiltonians and a variation of AM1 modified by the Bond Charge Correction procedure (BCC) [47], as well as a Density Functional Theory method which is considered in this work as the most precise assignment and therefore used as reference to compare the rest of the methods. Semiempirical models used in this work are based on the Neglect of Diatomic Differential Overlap (NDDO) method in which the overlap matrix S is replaced by the unit matrix. This allows one to replace the Hartree-Fock secular equation $|\mathbf{H}-\mathbf{E}S|=0$ with a simpler equation $|\mathbf{H}-\mathbf{E}|=0$. MNDO [48] is the oldest NDDO-based model that parameterizes one-center two-electron integrals based on spectroscopic data for isolated atoms, and evaluates other two-electron integrals using the idea of multipole-multipole interactions from classical electrostatics. AM1 [49] takes a similar approach to

MNDO in approximating two-electron integrals but uses a modified expression for nuclear-nuclear core repulsion. The modified expression results in non-physical attractive forces that mimic van der Waals interactions. The modification also required a model re-parameterization which was carried out with a particular emphasis on dipole moments, ionization potentials, and geometries of molecules. While this allows for some description of the hydrogen bond, other deficiencies, such as systematic over-estimates of basicities, remained. PM3 [50-51] uses a Hamiltonian that is very similar to the AM1 Hamiltonian but the parameterization strategy is different. While AM1 was parameterized based largely on a small number of atomic data, PM3 is parameterized to reproduce a large number of molecular properties. All DFT calculations were performed with Material Studio Modeling DMol³ software by Accelrys Inc. [52]. The non-local exchange and correlation functional BLYP [53-54] was used with a Double Numerical basis set including a polarization p-function for all hydrogens (a DNP basis set), which implies best accuracy at highest cost but important for hydrogen bonding [55]. The choice for the BLYP functional was guided by its good description of the atomic charges of different molecules [56-57]. Furthermore, it has been shown that DFT-BLYP gives a proper description of polar systems [58-59]. A fine integration grid was chosen for geometry optimizations convergence at high accuracy. The *integration accuracy* controls the precision with which Hamiltonian matrix elements are computed, as described in the DMol³ theory documentation. Partial charges were calculated from DFT results by Hirshfeld method [60] with the same software.

The Hirshfeld method (or ‘stockholder’ partitioning) uses the charge density distribution to determine atomic charges in the molecule. First, the reference state of the promolecule density is defined as $\rho^{pro}(\vec{r}) = \sum_A \rho_A(\vec{r})$, where $\rho_A(\vec{r})$ is the electron density of the isolated atom A placed at its position in the molecule. The atomic charge is

$$q_A = -\int \delta\rho_A(\vec{r}) d\vec{r}, \quad (1)$$

where $\delta\rho_A(\vec{r})$ is the atomic deformation density given by

$$\delta\rho_A(\vec{r}) = w_A(\vec{r}) \Delta\rho(\vec{r}). \quad (2)$$

In equation (2), $w_A(\vec{r})$ is the relative contribution (‘share’) of the atom A in the promolecule, whereas $\Delta\rho(\vec{r})$ is the molecular deformation density. The sharing factor is a weight that determines the relative contribution of atom x to the promolecule density at point r . It is defined as

$$w_A(\vec{r}) = \frac{\rho_A(\vec{r})}{\rho^{pro}(\vec{r})}. \quad (3)$$

Table 3. Number of the best scored docking results found on different orientations in the drug-site 1 of human albumin. A and A' represent positions depicted by arrow A in Figure 2 (xanthonoid moiety oriented to the interior of the pocket, in the direction of the arrow head). Primed and non-primed symbols correspond to a 180° turn along the arrow axis. B corresponds to position illustrated by arrow B in Fig. 2 xanthonoid moiety directed to the exterior of the moiety (arrow head). The symbol -B represent the same direction as B but xanthonoid directed to the interior and carbohydrate to the arrow head.

MGF	A	A'	B	-B	O	Range of scores
1 to 10	8	0	2	0	0	-5.98 to -5.26
11 to 20	5	0	2	2	1	-5.25 to -4.94
21 to 30	7	0	1	2	0	-4.92 to -4.69
31 to 40	6	0	2	2	0	-4.69 to -4.62
41 to 50	4	1	1	4	0	-4.61 to -4.51
TOTAL	30	1	8	10	1	
DP1						
1 to 10	8	0	0	2	0	-5.77 to -5.03
11 to 20	3	1	1	4	1	-5.03 to -4.84
21 to 30	3	1	2	1	3	-4.81 to -4.69
31 to 40	3	1	2	4	0	-4.68 to -4.58
41 to 50	3	0	1	5	1	-4.57 to -4.51
TOTAL	20	3	6	16	5	
DP6						
1 to 10	4	0	4	1	1	-6.35 to -5.32
11 to 20	1	1	6	2	0	-5.31 to -5.06
21 to 30	6	0	3	1	0	-5.05 to -4.94
31 to 40	1	2	4	2	1	-4.94 to -4.84
41 to 50	2	0	4	4	0	-4.83 to -4.66
TOTAL	14	3	21	10	2	

The molecular deformation density (used in equation (2)) is

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \rho^{pro}(\vec{r}), \quad (4)$$

where $\rho(\vec{r})$ is the molecular electron density. The Hirshfeld partitioning is almost insensitive to the basis set and it minimizes missing information [61-62].

All quantum mechanics calculations were done on the optimized geometry of neutral MGF obtained by DFT to eliminate conformational searches with quantum methods and to avoid introducing additional geometrical variables thus allowing better comparison of the results from different procedures. Conformational search methods are described in Results and Discussion since some of the parameters and procedures were not standard.

Docking procedures were performed using Human Serum Albumin (HSA) as receptor downloading its structure from the Protein Data Bank (PDB) [63]. Several different files of HSA complexed with different drugs and ligands are available from that web site. At least six different binding sites have been determined by crystallography for this enzyme. We selected a PDB file (ID code 2BXN) with the so called HSA drug-site 1

occupied by a polar and planar drug, iodipamide. The rationale of this selection is to choose the receptor with the ligand most similar to MGF and located in the most common drug binding site. Albumin is a 66 kDa monomer whose structure is formed by three homologous structural domains, each constituted by two subdomains named A and B. Drug-site 1 is found in subdomain 2A, formed by residues 193 to 299. For all docking studies performed only this subdomain was used. All hydrogen atoms were added and minimized with CHARMM27 force field and charge assignments maintaining heavy atoms in their crystallographic positions. Potential binding sites were identified with the alpha site method [64-65] and with CASTp server [66]. Only the binding site with the highest score was used in this work, and it is formed by atoms from 32 residues, a reflection of the site size. MOE docking studies used the alpha site triangle method to bias the orientation search of the ligand to meaningful trials. At least 60 thousand orientations were constructed and evaluated for each conformer, unless docking score converged in the last 15000 orientations. Using these parameters, iodipamide (in crystallographic conformation and with PEOE partial charges) was docked on drug-site 1 of albumin, obtaining a ligand orientation with an RMS deviation of just 0.39 Å with respect to the crystallographic structure.

When other conformers of iodipamide were used, the docking method detected the crystallographic conformation among the best five scores out of 200 energy minima. These results can be considered as a methodology validation. Docking studies of MGF were performed using charges determined by DFT since the main objective was to analyze the effect of speciation on docking results, and this method was considered the best available methodology for charge assignment. Nevertheless, the same docking studies with charges on MGF assigned by PEOE gave essentially the same results. This observation might be anticipated since DFT and PEOE charges are quite similar, with the exception of some oxygen atoms, and the main cause of changes in the ligand orientation is the existence and molecular localization of ionized groups.

The docking scoring-function used estimates the enthalpic contribution to the free energy of binding using a linear function:

$$\text{Score} = C_{\text{hb}}f_{\text{hb}} + C_{\text{ion}}f_{\text{ion}} + C_{\text{hh}}f_{\text{hh}} + C_{\text{hp}}f_{\text{hp}} + C_{\text{aa}}f_{\text{aa}}$$

Where the f represents different types of receptor-ligand contacts and the coefficients C weight the term contributions. The individual contacts are interactions between hydrogen bond donor-acceptor pairs (f_{hb}), coulombic ionic interactions (f_{ion}), hydrophobic interactions (f_{hh}), interactions between hydrophobic and polar atoms (f_{hp}) and van der Waals atom-atom contacts (f_{aa}).

Concluding Remarks

It was used mangiferin, a bioconjugate molecule with biological activity, as an archetype of carbohydrate-polyphenol potential drug lead to explore different important effects on virtual ligand design. Being xanthonoids along with flavonoids two important polyphenolic groups as drug leads and metabolites, our main contribution to drug design is to show that the ionization state of those polyphenolic compounds is difficult to predict and significant to consider for further computational studies. About methodologies to assign partial charges, it was found that PEOE is the best method for both sugar and phenolic moieties, with a good compromise between precision and computational cost. In addition, it was established the method capability to deal with neutral as well as anionic species in the xanthonoid rings. The charge assignment effect had demonstrated its relevance on conformational search, even with the relatively rigid structure of MGF. One of the effects is in the number of different structures, both before and after energy minimization, others are the difference in conformation obtained for the lower and lowest energy conformations and the magnitude in energy gaps between them. For docking simulations, speciation has proven its importance since choosing different species for the same molecule could lead to different results. In this case, two potential species showed a tendency to expose their carbohydrate region to solvent and to occupy the A region of the binding site in the potential receptor. In accordance with the speciation study, both neutral man-

giferin (MGF) and deprotonated in position 1 (DP1), are not the expected predominant ones at the physiological pH value. On the contrary, the properly assigned oxyanion (DP6) prefers to be allocated in the B region with its xanthonoid moiety exposed. Therefore, the present work has been able to exhibit the importance of experimental synergy with computational work and some of the complexity involved in modeling protein-ligand complexes.

Supporting information: Full atomic charge tables, molecular structures and coordinates from docking results are available from the authors upon request.

Acknowledgements

We are grateful for financial support received from National Council of Science and Technology (CONACyT, México) through grants 46168-M, 7242/040061 and ANUIES-ECOS Nord M05-S01. Also, we thank for support grants from UAM, México (Acuerdos 11/07 y 13/07 del Rector General de la UAM: Programa de Fomento al Desarrollo de Grupos de Académicos en Formación y Programa de Apoyo a la Investigación Multidisciplinaria, cuentas 8110117 y 8110118).

References

- Reddy R.N.; Mutyala, R.; Aparoy, P.; Reddanna, P.; Reddy, M. R. *Curr. Pharm. Des.* **2007**, 13, 3505-3517.
- Van Drie, J.H. *J. Comput. Aided Mol. Des.* **2007**, 21, 591-601.
- Geromichalos, G. D. *J. BUON.* **2007**, Suppl 1, S101-S118.
- Zeinalipour-Loizidou, E.; Nicolaou, C.; Nicolaides, A.; Kostrikis, L. G. *Curr HIV Res.* **2007**, 5, 365-388.
- Dubinina, G. G.; Chupryna, O.O.; Platonov, M.O.; Borisko, P.O.; Ostrovska, G.V.; Tolmachov, A. O.; Shtil, A. A. *Anticancer Agents Med. Chem.* **2007**, 7, 171-188.
- Reddy, A.S.; Pati, S.P.; Kumar, P.P.; Pradeep, H.N.; Sastry, G. N. *Curr. Protein Pept. Sci.* **2007**, 8, 329-351.
- Kishan, K. V. *Curr. Protein Pept. Sci.* **2007**, 8, 376-380.
- Kroemer, R. T. *Curr. Protein Pept. Sci.* **2007**, 8, 312-328.
- Cavasotto, C. N.; Orry, A. J. *Curr. Top. Med. Chem.* **2007**, 7, 1006-1014.
- Joseph-McCarthy, D.; Baber, J. C.; Feyfant, E.; Thompson, D. C.; Humblet, C. *Curr. Opin. Drug. Discov. Devel.* **2007**, 10, 264-274.
- Gómez-Zaleta, B.; Ramírez-Silva, M. T.; Gutiérrez, A.; González-Vergara, E.; Güizado-Rodríguez, M.; Rojas-Hernández, A. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2006**, 64, 1002-1009.
- Templeton, D. M.; Ariese, F.; Cornelis, R.; Danielsson, L. G.; Muntau, H.; van Leeuwen, H. P.; Lovinski, R. *Pure Appl. Chem.* **2000**, 72, 1453-1470.
- Fotie, J.; Bohle, D. S. *Anti-Infective Agents in Medicinal Chemistry.* **2006**, 5, 15-31.
- Roberts, J. C. *Chem. Rev.* **1961**, 61, 591-605.
- Lin, C. N.; Chang, C. H.; Arisawa, M.; Shimizu, M.; Morita, N. *Phytochemistry.* **1982**, 21, 205-208.
- Gerassim, M.K.; Paraskev, T.N. *Biochem. Systemat. Ecol.* **1998**, 26, 647-653.
- McKay, D. L.; Blumberg, J. B. *Phytother. Res.* **2007**, 21, 1-16.
- Garrido, G.; González, D.; Delporte, C.; Backhouse, N.; Quintero, G.; Núñez-Sellés, A. J.; Morales, M. A. *Phytother Res.* **2001**, 15, 18-21.

19. Ichiki, H.; Miura, T.; Kubo, M.; Ishihara, E.; Komatsu, Y.; Tanigawa, K.; Okada, M. *Bio. Pharm. Bull.* **1998**, *21*, 1389-1390.
20. Miura, T.; Ichiki, H.; Hashimoto, I.; Iwamoto, N.; Kato, M.; Kubo, M.; Ishihara, E.; Komatsu, Y.; Okada, M.; Ishida, T.; Tanigawa, K. *Phytomedicine*. **2001**, *8*, 85-87.
21. Iwamoto, N.; Kato, M.; Ichiki, H.; Kubo, M.; Komatsu, Y.; Ishida, T.; Okada, M.; Tanigawa, K. *Biol. Pharm. Bull.* **2001**, *24*, 1091-1092.
22. Muruganandan, S.; Srinivasan, K.; Gupta, S.; Gupta, P. K.; Lal, J. *J. Ethnopharmacol.* **2005**, *97*, 497-501.
23. Yoshimi, N.; Matsunaga, K.; Katayama, M.; Yamada, Y.; Kuno, T.; Qiao, Z.; Hara, A.; Yamahara, J.; Mori, H. *Cancer Lett.* **2001**, *163*, 163-170.
24. Guha, S.; Ghosal, S.; Chattopadhyay, U. *Chemotherapy*. **1996**, *42*, 443-451.
25. Ghosal, S.; Chaudhuri, R. K. *J. Pharm. Sci.* **1975**, *64*, 888-889.
26. Miliuskas, G.; Venskutonis, P. R.; van Beek, T. A. *Food Chem.* **2004**, *85*, 231-237.
27. Murunganandan, S.; Lal, J.; Gupta, P. K.; *Toxicology*. **2005**, *215*, 57-68.
28. Ghosal, S.; Rao, G.; Saravanan, V.; Misra, S.; Rana, D. *Indian J. Chem. B. Org.* **1996**, *35B*, 561-566.
29. Zhang, H. Y. *Acta Pharmacol. Sin.* **1999**, *20*, 555-558.
30. Gasteiger, J.; Marsili, M. *Tetrahedron*, **1980**, *36*, 3219-3228.
31. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
32. Gomez-Zaleta, B. Ph.D. Thesis. Universidad Autónoma Metropolitana-Iztapalapa, (2006). <http://148.206.53.231/UAMI13182.pdf>, accessed in October, 2007.
33. Advanced Chemistry Development, Inc. *ACD Labs*. Toronto ON, Canada, 2003. <http://www.acdlabs.com>, accessed in October, 2007.
34. Carter, D. C.; Ho, J. X. *Adv. Protein Chem.* **1994**, *45*, 153-203.
35. Zsila, F.; Bikadi, Z.; Simonyi, M. *Biochem. Pharmacol.* **2003**, *65*, 447-456.
36. Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. *J. Mol. Biol.* **2005**, *353*, 38-52.
37. Núñez-Sellés, A. J.; Velez, H. T.; Agüero-Agüero, J.; González-González J.; Naddeo, F.; De Simone, F.; Rastelli, L. *J. Agric. Food Chem.* **2002**, *50*, 762-769.
38. Chemical Computing Group, Inc. *Molecular Operating Environment (MOE)*. CCG, Montreal, Canada. 2007. <http://www.chemcomp.com>, accessed in February, 2008.
39. Halgren, T. A. *J. Comput. Chem.* **20**, **1999**, 720-729.
40. Halgren, T. A. *J. Comput. Chem.* **20**, **1999**, 730-748.
41. MacKerell A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L. Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B.* **1998**, *102*, 3586-3616.
42. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400-1415.
43. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
44. Fabricius, J.; Engelsens, S. B.; Rasmussen, K. *J. Carbohydr. Chem.* **1997**, *16*, 751-772.
45. Clark, M.; Cramer, R. D., III; van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982-1012.
46. Engh, R. A.; Huber, R. *Acta Crystallogr. A.* **1991**, *47*, 392-400.
47. Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
48. Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899-4907.
49. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
50. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209-220.
51. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221-264.
52. Delley, B. *Materials Studio DMol³ version 3.0. Density Functional Theory Electronic Structure Program*. Accelrys Inc. 2003.
53. Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547-2553.
54. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785-792.
55. Delley, B. *J. Chem. Phys.* **1990**, *92*, 508-517.
56. Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, **2000**.
57. Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comp. Chem.* **2003**, *24*, 1291-1304.
58. Choi, C. H.; Kertesz, M. *Chem. Phys. Lett.* **1996**, *263*, 697-702.
59. Heiberg, H.; Gropen, O.; Laerdahl, J. K.; Swang, O.; Wahlgren, U. *Theor. Chem. Acc.* **2003**, *110*, 118-125.
60. Hirshfeld, F. L. *Theor. Chim. Acta B*, **1977**, *44*, 129-138.
61. De Proft, F.; Van Alsenoy, C.; Peeters, A.; Langenaeker, W.; Geerlings, P. *J. Comput. Chem.* **2002**, *23*, 1198-1209.
62. Nalewajski, R. F.; Parr, R. G. *Proc. Natl. Acad. Sci. USA.* **2000**, *97*, 8879-8882.
63. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
64. Edelsbrunner, H.; Facello, M.; Fu, R.; Liang, J. *Proceedings of the 28th Hawaii International Conference on Systems Science.* **1995**, 256-264.
65. Liang, J.; Edelsbrunner, H.; Fu, P.; Sudhakar, P. V.; Subramaniam, S. *Proteins.* **1998**, *33*, 1-17.
66. Liang, J.; Edelsbrunner, H.; Woodward, C. *Protein Sci.* **1998**, *7*, 1884-1897.